

Miquel TérmensDepartament de Biblioteconomia
i Documentació, Universitat
de Barcelona i Institut d'Estudis
Catalans Studies, Barcelona

El DOI: el «Gran Germà» de la comunicació científica

Adreça per a la correspondència:
E-mail: termens@ub.edu

Els identificadors digitals

L'augment de la disponibilitat i l'ús de documentació en suport digital ha dut aparellat la necessitat de desenvolupar mecanismes que en facilitin la manipulació. Una de les eines clau per a fer-ho possible són els identificadors digitals. Entenem per *identificadors digitals* els sistemes de codificació única que permeten individualitzar i localitzar els objectes digitals, normalment al nivell de fitxer informàtic. Es tracta de sistemes estandarditzats a nivell global, però aplicats a sectors o productes determinats. Els identificadors digitals s'engloben dins del ja elevat nombre d'identificadors desenvolupats per a facilitar la manipulació de grans volums d'objectes i informacions des d'un punt de vista de transacció i, sovint, amb un abast mundial. Podem destacar, per l'àmplia acceptació que té, el número d'article de comerç global (GTIN, '*global trade item number*'), que és el codi que identifica de manera inequívoca els articles comercials arreu del món. El GTIN pot adoptar formes diferents, segons la finalitat que tingui: EAN-12, EAN-8, EAN-14 o UCC-12. L'EAN-12 és el codi utilitzat en els productes finals de distribució, amb l'objectiu de facilitar-ne la comercialització en el punt de venda als clients minoristes; el seu símbol o forma gràfica, que tots coneixem, són els codis de barres EAN/UPC-13 [1].

En l'àmbit de la documentació, prèviament a l'aparició dels documents digitals, s'havien desenvolupat identificadors o codis que facilitaven la gestió de la informació [2]. En

podem destacar l'ISBN, l'ISSN i el número del Dipòsit Legal. El número internacional normalitzat per als llibres (ISBN, '*international standard book number*') és un codi mundial que s'assigna a tots els llibres publicats; actualment cobreix dues necessitats bàsiques: facilitar el comerç editorial (catàlegs de les editorials, comandes de les llibreries) i facilitar el control bibliogràfic que n'han de fer les biblioteques; la gestió d'aquesta numeració està coordinada internacionalment per The International ISBN Agency [3] i es desenvolupa en cada país mitjançant l'agència nacional corresponent, que depèn de l'Estat de manera directa o bé és una concessió privada.

El número internacional normalitzat per a publicacions en sèrie (ISSN, '*international standard serial number*') ha estat concebut paral·lelament a l'ISBN i funciona d'una manera similar, per bé que s'aplica a les publicacions en sèrie, especialment les revistes; és gestionat per l'ISSN International Center [4], dependent de la UNESCO, i per agències estatals. El número del Dipòsit Legal no és un sistema universal, sinó que és propi de determinats països i és regulat segons la seva pròpia legislació; l'objectiu inicial era facilitar a l'Administració el control sobre els editors de llibres pel que fa al compliment de les normes de propietat intel·lectual i, indirectament, exercir la censura d'impresament, però actualment l'ús predominant és el d'instrument pel qual l'Estat s'assegura l'obtenció d'algun exemplar de totes les obres publicades al seu territori, que després es destinarà a la biblioteca nacional de cada país.

L'extensió de la documentació en suport digital va obligar a trobar noves solucions. Això es va deure, d'una banda, a la inadequació de l'ús dels identificadors existents —que havien estat pensats per a usar-los en objectes físics tangibles— en els objectes digitals intangibles, i, per una altra banda, a les característiques mateixes del mitjà electrònic, com són les noves formes d'ús de la documentació o les possibilitats de manipulació informàtica dels identificadors. Les solucions aportades es poden aplegar en tres categories:

a) Desenvolupaments a partir dels codis preexistents; el cas més reeixit és el de l'identificador de contribució i número de publicació en sèrie (SICI, '*serial item and contribution identifier*'), aplicable, per exemple, als articles de revista [5].

b) Creació de codis nous per a nous objectes digitals, com el codi normalitzat internacional d'enregistrament (ISRC, '*international standard recording code*'), per a música gravada [6].

c) Creació de nous sistemes globals d'identificació; dins d'aquesta categoria trobem el DOI.

El DOI

L'identificador d'objecte digital (DOI, '*digital object identifier*') [7] és un codi identificador que té les propietats següents: s'aplica bàsicament als objectes digitals, no se centra en un única categoria de documentació, té un caràcter mundial amb un únic sistema de numeració, té una gestió de caràcter altament centralitzat, la governació i el manteniment no són a càrrec d'entitats governamentals sinó d'entitats dependents del món privat, i té diverses aplicacions operatives. Breument, es pot definir el DOI com un identificador que, de manera voluntària, poden incorporar els documents publicats, de qualsevol naturalesa i suport, amb la finalitat de facilitar-ne la gestió automatitzada. La voluntat del DOI no és substituir els sistemes anteriors —com ara l'ISBN, l'ISSN i el SICI—, sinó cobrir els buits que deixaven i adaptar totes les seves funcionalitats a la gestió automatitzada i als documents digitals.

Des d'un punt de vista tècnic, el sistema està format per quatre components:

a) Una sintaxi de numeració estandarditzada. S'assigna un identificador únic a cada document, amb el nivell de granularitat que decideixi l'editorial. El codi està format per dues parts, un prefix i un sufix, separades per una barra. El prefix conté el nombre que identifica l'editorial del document. El sufix identifica el document concret entre els produïts per una editorial determinada; el sufix és assignat per la mateixa editorial i no té una forma normalitzada, però es recomana

l'ús de codis acceptats en el camp corresponent, com l'ISBN, en el cas dels llibres, o el SICI, en el cas dels articles de revista.

b) Un sistema de resolució, basat en el sistema identificador anomenat *The Handle System*, creat per la Corporació d'Iniciatives Nacionals de Recerca (Corporation for National Research Initiatives) dels EUA. El sistema permet obtenir la URL actual d'un document a partir del seu codi DOI o de les metadades que el descriuen.

c) Un model i un diccionari de dades. Permet usar diferents esquemes de metadades per a descriure els documents.

d) Un mecanisme d'implementació basat, des d'un punt de vista organitzatiu, en la Fundació Internacional del DOI i per les agències de registre del DOI, i, des del punt de vista tècnic, en un conjunt de bases de dades interrelacionades.

El DOI permet identificar, descriure i resoldre la localització d'un recurs digital. En aquest sentit, dona més informació que la major part d'identificadors, els quals es limiten, com el seu nom indica, a facilitar una identificació correcta de l'objecte a què es refereixen. En el cas del DOI, quan una editorial assigna un DOI, també està obligada a introduir en la base de dades central les metadades que descriuen el document i que poden servir a terceres parts per a fer-ne referència. A més, ha d'indicar com és la URL operativa del document i mantenir aquesta dada actualitzada; amb això el sistema de resolució DOI pot facilitar a tercers la localització a Internet del document a partir de la introducció del seu codi DOI.

Com ja s'ha indicat, el sistema és gestionat per la Fundació Internacional del DOI (IDF, '*International DOI Foundation*'), creada el 1998, una entitat sense ànim de lucre fundada per uns quants dels grups editorials més importants. La Fundació s'encarrega dels aspectes tècnics, però no de les tasques de registre, que estan delegades en les agències de registre del DOI (DRA, '*DOI Registration Agencies*'). Les DRA són entitats independents, de caràcter públic o privat, que han obtingut l'autorització per a assignar prefixos a les editorials que ho demanin i a gestionar els codis DOI que aquests assignin. Cal destacar diverses característiques en aquest model: les DRA no són agències de caràcter estatal, com sí que ho són les que gestionen l'ISBN i l'ISSN, i n'hi haurà un nombre molt limitat, amb la qual cosa es configurarà un sistema altament centralitzat en unes poques agències; a conseqüència d'aquest fet, el funcionament de les DRA no és subvencionat pels estats, i els seus serveis s'han de pagar; finalment, les DRA no solament ofereixen el servei de registre sinó també els serveis que hi estan associats (accés a *The Handle System*, etc.). En el moment de redactar aquest treball, solament hi ha set agències autoritzades [8].

Les principals aplicacions del DOI són fer enllaçaments de referències, generar URL persistents, gestionar drets de

propietat intel·lectual i crear referències bibliogràfiques. Cadascuna d'aquestes aplicacions acaba tenint una conformació concreta segons el tipus de document o la finalitat amb què s'utilitza. Per això les DRA, atès que no es diferencien per demarcacions territorials, s'han especialitzat a atendre determinades especialitats editorials i els serveis específics que aquestes requereixen; així, CrossRef se centra a gestionar els codis DOI i els serveis associats corresponents a les revistes científiques.

CrossRef

CrossRef va ser la primera agència que es va autoritzar; de fet, els seus objectius van ser la raó d'ésser del sistema DOI. A l'inici de l'any 2000 es va crear l'organització sense ànim de lucre Publishers International Linking Association, Inc. (PILA), operadora del sistema CrossRef. Actualment, la direcció està formada per representants d'AAAS (Science), AIP, ACM, APA, Blackwell Publishers, Elsevier Science, IEEE, Wolters Kluwer, Nature, University of Chicago Press, Sage, Springer, Taylor & Francis, Thieme i Wiley. Les editorials afiliades són més de mil quatre-centes, i es pot considerar que aquests conformen un percentatge molt elevat de la publicació de revistes científiques, especialment, les de caràcter comercial i abast mundial. Dit d'una altra manera, CrossRef és la plataforma creada per les principals editorials de revistes científiques digitals per a assignar codis DOI i, gràcies a aquests codis, poder automatitzar la tasca d'enllaçar amb hipertext les referències bibliogràfiques que es donen entre les unes i les altres [9]. Aquesta activitat, coneguda tècnicament per *enllaçament de referències* (en anglès, *reference linking*), no era abordable de manera manual, atès el gran volum de minucios treball manual que requeria: buscar les notes i les referències finals que conté un article, centrar-se en les que contenen referències bibliogràfiques a revistes, interpretar-les (cal recordar els múltiples sistemes de citació que han imposat les diferents revistes), comprovar que la revista citada estigui disponible en versió electrònica, consultar la referència concreta (volum, número, article), anotar la URL i, finalment, crear en l'article origen un hiperenllaç amb la URL localitzada.

CrossRef simplifica i automatitza aquest procés. Les editorials associades estan obligades a assignar codis DOI als seus articles, a emplenar amb el sistema DOI (operació que es pot automatitzar) una fitxa amb les metadades de cadascun d'ells (autor/a, títol, revista, etc.) i la URL operativa, i, finalment, a mantenir actualitzada l'adreça URL, si aquesta varia. A canvi, les editorials poden consultar de manera automatitzada la base de dades central; en el cas de les edito-

rials grans i mitjanes, un programa s'encarrega de generar els enllaços: rastreja el text dels nous articles; extreu les referències a altres articles, les interpreta i genera una recerca en el The Handle System; en aquest, els termes de recerca són comparats amb les seves metadades, que apunten a un codi DOI i aquest a una URL operativa; el resultat obtingut s'escriu en l'article d'origen. The Handle System també ofereix una altra millora: les URL són sempre operatives i no requereixen cap manteniment. L'obsolescència de les URL, amb el resultat d'una gran quantitat d'enllaços trencats, és un dels majors problemes que presenta l'arquitectura actual d'Internet i que frena l'establiment de més enllaços entre llocs web [10]. The Handle System ofereix una solució, ja que crea, per a cada document, una adreça permanent —un nom de recurs uniforme (URN, '*uniform resource name*')— que pot ser utilitzada per a localitzar-lo. Els codis DOI poden actuar com a URN si se'ls anteposa l'adreça del servidor de resolució (per exemple: <http://dx.doi.org/10.1126/science.1088234>); d'aquesta manera, els lectors no s'han de preocupar per la vigència dels enllaços, ja que d'això s'encarrega l'editorial de la revista citada, que manté la URL correcta en les bases de dades de DOI.

Altres aplicacions

Cal destacar que l'ús de Crossref i els dos serveis que comporta —enllaçament de referències i URLs persistents— han estat dues de les claus que expliquen el gran desenvolupament de les revistes científiques electròniques i la ràpida acceptació que han tingut dels lectors. Les aplicacions del DOI no acaben aquí, encara que aquests siguin els seus resultats més evidents. El sistema també es pot encarregar de controlar els drets d'accés i ús dels usuaris als documents, així com mantenir una estadística centralitzada de l'ús que se'n fa; en un sentit més ampli, ja s'ofereixen serveis orientats al control dels drets de propietat intel·lectual, incloent-hi la gestió econòmica. Altres aplicacions estan en desenvolupament; per exemple, la Biblioteca d'Informació Tècnica d'Alemanya (TIB, '*Technische Informationsbibliothek*') està experimentant l'assignació del DOI a dades científiques primàries, com poden ser taules climàtiques i imatges obtingudes per satèl·lit [11].

En qualsevol cas, sembla evident que l'extensió del DOI anirà molt lligada als interessos i les necessitats de la indústria editorial, que al cap i a la fi és qui el va crear. En aquest sentit, es pot observar que s'han obtingut grans resultats pel que fa a interconnexió i a aplicacions corporatives dins d'aquest sector; però els resultats han estat més aviat minsos en relació amb altres actors interessats, com ho poden ser les

petites editorials o les biblioteques. Les editorials petites o sense ànim de lucre s'han trobat amb uns requeriments tècnics i econòmics (recordem que tots els serveis del DOI són de pagament) que han frenat l'adopció del DOI com valor afegit de les seves publicacions. Les biblioteques, per la seva banda, s'han trobat amb el problema de la còpia adient: el sistema DOI ofereix la URL del document objecte de referència, però en alguns casos aquest es troba dins d'un servei diferent del que la biblioteca ha contractat; en definitiva, els enllaços DOI són propis d'un sistema tancat i no són sensibles al context. Per a remeiar-ho, s'han desenvolupat solucions com l'estàndard OpenURL i el sistema de resolució SFX [12].

Per acabar, s'han de destacar uns altres dos fenòmens. El DOI va néixer i s'ha desenvolupat a l'empara de les necessitats de l'edició de revistes electròniques, i pot ser que s'acabi aplicant a altres activitats documentals, com els sistemes d'aprenentatge virtual. En canvi, de moment no està tenint gaire incidència entre els sistemes de llibres electrònics, potser perquè aquest mercat està trobant dificultats per a desenvolupar-se. Finalment, hi ha qui recela de les noves aplicacions que puguin aparèixer en el futur, recordant que, per primera vegada en la història, els usos que la comunitat científica internacional està fent de les revistes científiques (cada consulta que fem a un article científic) estan quedant registrats en un sistema centralitzat de bases de dades, sostingut, a més, de manera privada per les empreses comercials del sector. Aquesta arquitectura està facilitant la realització d'anàlisis d'enllaços citats que permeten conèixer millor l'ús que la comunitat científica dóna a les publicacions disponibles —com ho indica la incorporació progressiva que n'ha fet l'Institut per a la Informació Científica (ISI, 'Institut for Scientific Information')—; però també es podria convertir en un instrument per a modificar les pautes de funcionament de l'edició científica.

Notes

- [1] L'organisme encarregat de normalitzar-lo és el GS1, <http://www.ean-int.org>.
- [2] G. VITIELLO (2004), «Identifiers and identification systems: an informational look at policies and roles from a library perspective», *D-Lib Magazine*, vol. 10, núm. 1 (gener), <<http://www.dlib.org/dlib/january04/vitiello/01vitiello.html>>
- [3] <http://www.isbn-international.org>.
- [4] <http://www.issn.org>.
- [5] <http://www.niso.org/standards/resources/Z39-56.pdf>.
- [6] <http://www.ifpi.org/isrc>.
- [7] <http://www.doi.org>.
- [8] http://www.doi.org/registration_agencies.html
- [9] A. BRAND (2001), «CrossRef turns one». *D-Lib Magazine*, vol. 7, núm. 5 (maig) (2002), «Metadata and reference linking».
- [10] Un estudi recent sobre més de mil articles publicats entre el 2000 i el 2003 en les revistes *New England Journal of Medicine*, *The Journal of the American Medical Association* i *Science*, va demostrar que el 13 % de les referències d'Internet que citaven estaven inactives [R. P. DELLAVALLE (2003), «Going, going, gone: lost internet references», *Science*, vol. 302, núm. 5646 (octubre), p. 787-788]. Un altre estudi anterior indicava que la vida mitjana de les URL en articles d'informàtica era de quatre anys a partir de la publicació [D. SPINELLIS (2003), «The decay and failures of URL references», *Communications of the ACM*, vol. 46, núm. 1 (gener), p. 71-77, <<http://www.spinellis.gr/sw/url-decay/>>].
- [11] N. PASKIN (2005), «Digital object identifiers for scientific data», *Data Science Journal*, vol. 4, p. 12-20, <http://www.jstage.jst.go.jp/article/dsj/4/0/12/_pdf>.
- [12] H. van de SOMPEL i O. BEIT-ARIE (2001), «Open linking in the scholarly information environment using the OpenURL framework». *D-Lib Magazine*, vol. 7, núm. 3 (març), <<http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html>>.